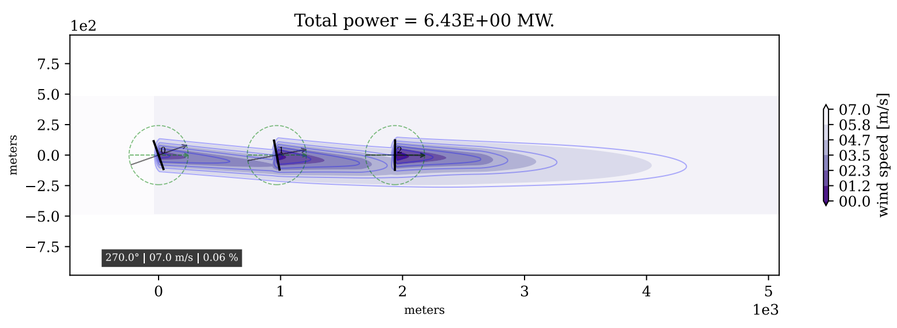
Helping mitigate climate change through efficient reinforcement learning-based wind farm flow control

Elie Kadoche[1,2], Pascal Bianchi[1], Florence Carton[2], Philippe Ciblat[1], Damien Ernst[1,3]

[1]Polytechnic Institute of Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France
[2]TotalEnergies OneTech, 2 Place Jean Millier, 92400 Courbevoie, France
[3]Montefiore Institute, University of Liège, 4000 Liège, Belgium

1) Wind farm flow control

WAKE STEERING



PROBLEM

$$\max_{\mathbf{a}_t, \mathbf{a}_{t+1}, \ldots, \mathbf{a}_{t+L-1}} \sum_{k=0}^{L-1} \sum_{i=0}^{N-1} \mathop{\mathbb{E}}_{W \sim D} \left[P_{t+k}^i(\ldots) \right]$$

$\mathbf{a}_{t+k} = (a_{t+k}^0, a_{t+k}^1, \ldots)$ yaw settings

$a_{\min} \leqslant a_{t+k}^i \leqslant a_{\max}$ rotational constraints

$W \sim D$ stochastic wind

$P_{t+k}^i(\ldots)$ power output

METHODS

- Myopic control.
- Predictive control.
- Reinforcement learning.

$$\max_{\theta} \mathop{\mathbb{E}}_{\pi_\theta} \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \right]$$

$$\pi_\theta(s_t) = (\bar{\mu}_t, \bar{\kappa}_t, v_t)$$

$$a_t^i \sim \text{von Mises}(\mu_t^i, \kappa_t^i)$$

2) Generalization

ISSUE $\longrightarrow$ current RL methods only focus on limited sets of wind conditions.
CAUSE $\longrightarrow$ bad reward shaping (unbalanced across wind conditions).
SOLUTION $\longrightarrow$ better reward shaping (balanced across wind conditions).

$$r_{t+1} = \frac{1}{N} \sum_{i=0}^{N-1} \frac{P_t^i}{\Phi(V_t)}$$

Traditional reward.

When wake steering is optimal ...
Optimal reward: close to 0.5.

When wind tracking is optimal ...
Optimal reward: close to 1.

The policy learns a wind tracking solution, and it is enough to get high rewards.

$$\Delta_{P_t} = (P_t - \bar{P}_t)/\bar{P}_t$$

$$r_{t+1} = \Delta_{P_t} \mathbf{1}_{\Delta_{P_t} < 0} + \exp(-p\bar{\mathcal{L}}_t) \Delta_{P_t} \mathbf{1}_{\Delta_{P_t} \geq 0}$$

penalize policies below baseline

restrict Δ_{P_t} when it is supposed to be high (strong wake conditions)

encourage policies above baseline

More balanced reward (our contribution).

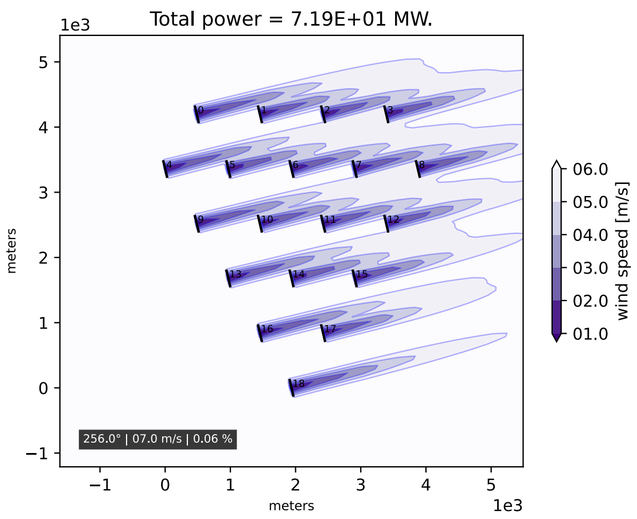
3) Sample efficiency

ISSUE $\longrightarrow$ training a model requires a large number of simulations.
CAUSE $\longrightarrow$ fully-connected neural network architecture.
SOLUTION $\longrightarrow$ attention-based neural network architecture.

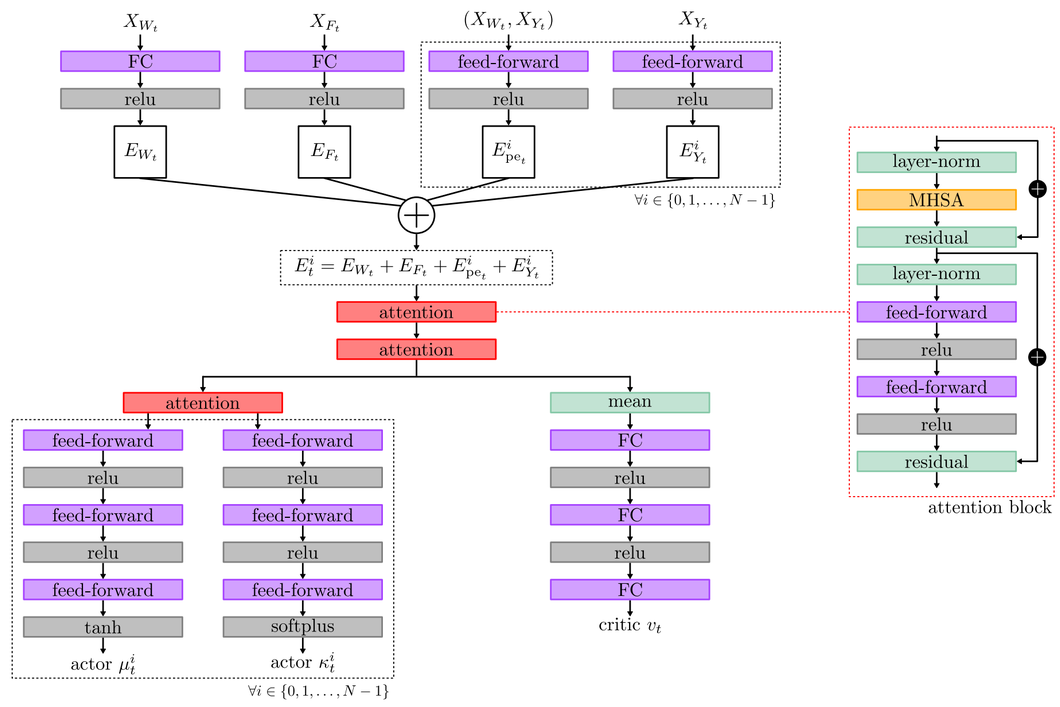
STATE

$$s_t = (X_{W_t}, X_{F_t}, X_{Y_t})$$

- X_{W_t} current wind data (noisy).
- X_{F_t} wind forecast (noisy).
- X_{Y_t} yaw angles.

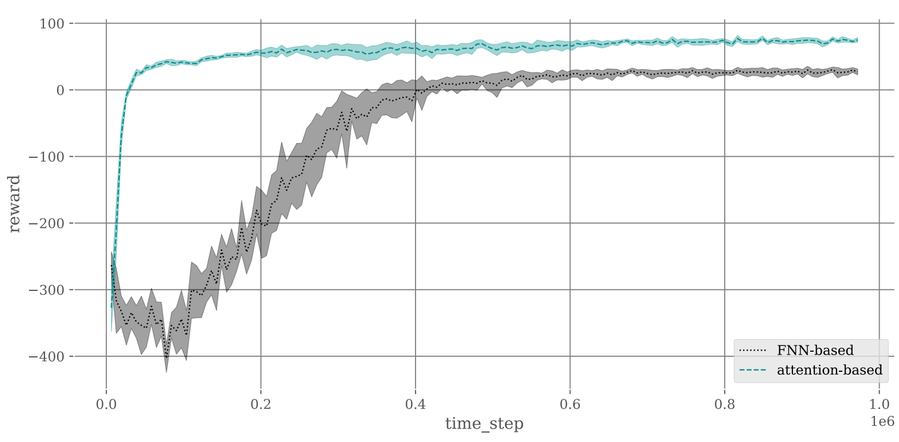


Wind farm.

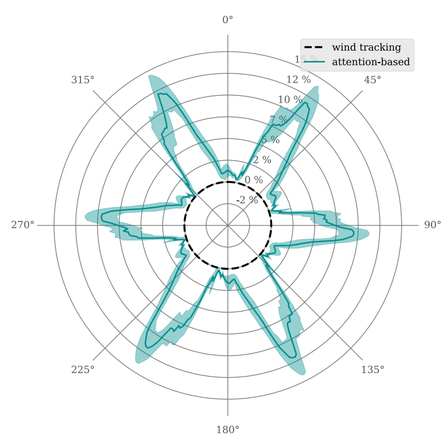


Attention-based model (our contribution).

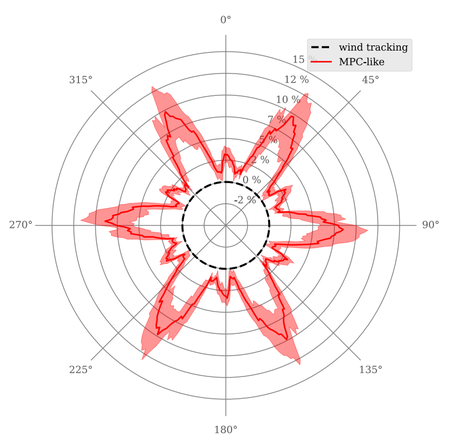
4) Numerical simulations



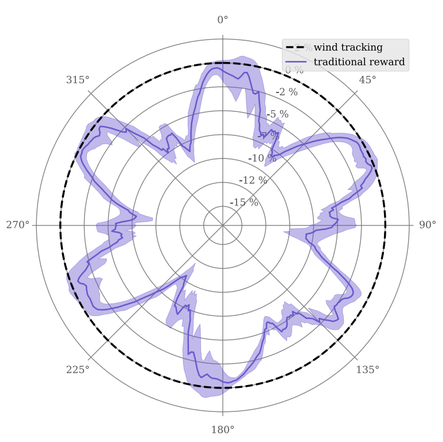
Training each model for 10 different seeds.



Testing: attention-based model.



Testing: MPC-like solution.



Testing: traditional reward.

* The attention-based model requires $\approx$ 10 times fewer simulation steps than the FNN-based model and it reaches superior performance.
* The attention-based model achieves performance comparable to the MPC-like solution, but with lower variance and higher gains.
* When using the traditional reward, the attention-based model completely fails to learn a wake steering strategy.